DOCUMENT RESUME

ED 244 976 TM 840 280

AUTHOR Hand, Carol A.; And Others
TITLE The Influence of Course upon Measures of Teaching Effectiveness in Mathematics.
PUB DATE Apr 84
NOTE 14p.; Paper presented at the Annual Meeting of the Mathematical Association of America, Southeastern Section (63rd, Cookeville, TN, April 6-7, 1984). Small print in Tables 1 and 2.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Advanced Courses; Analysis of Variance; *Courses; Higher Education; Introductory Courses; Majors (Students); *Mathematics Instruction; Nonmajors; Student Attitudes; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Teacher Evaluation; Test Reliability; *Test Validity

ABSTRACT

A frequently used method of measuring teaching effectiveness is the anonymous student rating of the instructor at the end of a grading period. The validity of these ratings for faculty personnel decisions has been a source of controversy. The purposes of this study were to examine the differences in teaching effectiveness between selected courses; investigate the effects of course type and course level on measures of teaching effectiveness, as well as possible interactions between type and level; and examine the differences between measures of effectiveness and their reliability. Data consisted of students' evaluations of undergraduate and graduate mathematics course instructors. Analysis of variance was used to compare the mean ratings from 20 mathematics courses in the study. Results indicated that the particular course, its type, and its level are important factors to consider when using student evaluations of teaching performance. (DWH)

The Influence of Course Upon Measures of Teaching Effectiveness in Mathematics

Carol A. Hand, Research Analyst
Office of Institutional Planning
Georgia State University
University Plaza
Atlanta, Georgia 30303 (404) 658-2570

Fred A. Massey, Chairman
Department of Mathematics
Georgia State University

James E. Prather, Senior Research Associate
Office of Institutional Planning
Georgia State University

Paper Presented to 63rd Annual Meeting
Mathematical Association of America, Southeastern Section
Tennessee Technological University
Cookeville, Tennessee
April 6 - April 7, 1984

The Influence of Course Upon Measures of Teaching Effectiveness in Mathematics

Research on evaluations of classroom teaching has produced various measures of teaching effectiveness. Self evaluation and peer evaluation have been used occasionally, with less than clear results. A more often used method of measuring teaching effectiveness is with anonymous student ratings of the instructor at the end of a grading period.

A controversy exists in the literature concerning the validity of using students' ratings for faculty personnel decisions. The findings of Marsh (1982) demonstrated agreement between students and instructors on evaluations of teaching effectiveness, and support the validity of student ratings. Dowell and Neal (1982) provide a review of studies which have attempted to link student ratings to student learning as a way of validating student ratings as a measure of teaching ability. The validity of student ratings is quite variable, and is at best only modest. They recommend that student ratings be used with great caution in the processes of faculty review and decision making. Hills (1974) concluded that student ratings of faculty could not be trusted when determining pay increases, promotion, and tenure.

Numerous studies have been conducted to investigate the factors that may influence or bias students' ratings. Hoffman and Kremer (1980) found student attitude and instructor attitude as perceived by the student to be important variables in predicting student ratings of the instructor. Personality characteristics of the instructor have also been shown to influence evaluations (Braskamp, Ory, and Pieper, 1981; Abrami, Perry, and Leventhal, 1982). The relationship between grades and instructor ratings has often been addressed. Several authors (Abrami, Dickens, Perry, and

Leventhal, 1980; Ducette and Kenny, 1982) have found grades to have significant effect on ratings, while others (Howard and Maxwell, 1980, 1982) argue against a grading leniency bias model.

Other studies focus on how student and course variables relate to students' evaluations of teaching. Marsh (1980) examined the relationship between student evaluations and certain background characteristics. Favorable student ratings were correlated with prior subject interest, higher expected grades, higher levels of workload difficulty, and a higher percentage of students taking the course for general interest only. Overall and Marsh (1980) investigated the relative contribution of course level (undergraduate versus graduate), course type (accounting, economics, etc.), and the specific instructor on students' evaluation. The variance which could be attributed to the specific instructor was much greater than that due to course level or course type. Who teaches a course appeared to be relatively more important than the particular course or the level at which it is taught.

Greene, Prather, and Sturgeon (1983) have introduced a unique and unobtrusive measure of teaching effectiveness. It is based on observable student behavior, and makes use of existing administrative data. This measure is the number of times students return to a particular teacher for additional courses. There is evidence that this measure of students' repeating faculty members can be a valid indicator of teaching effectiveness. Prather, Massey, and Greene (1983) found the repeat measure clearly related to students' ratings of instructors in introductory statistics courses. Students repeating a given faculty member was also found to be associated

with higher students' evaluations of instructors in mathematics courses (Prather, Massey, Greene, and Sturgeon, 1984).

This study focuses upon several measures of teaching effectiveness in mathematics courses. These are seven items of a teaching performance scale as well as the previously discussed unobtrusive measure of students repeating an irstructor.

The purposes of the study are 1) to examine differences in teaching effectiveness between selected courses, 2) to investigate the effects of course type and course level on measures of teaching effectiveness, as well as possible interactions between type and level, and 3) to look at differences among the measures of effectivenss and at the reliability of such measures.

Method

Data

The data consist of students' evaluations of undergraduate and graduate mathematics course instructors for the period 1979 to 1982. A total of 20 courses, 590 classes, and 9144 evaluations was considered. An example of the evaluation instrument and the way it is scored can be found in the Appendix. Both service courses and courses for mathematics majors were included.

Procedure

Analysis of variance is used to compare the mean ratings from the 20 individual mathematics courses. The independent variables are the course (Intermediate Algebra, Calculus I, etc.), the course type (service versus for degree majors), and course level (freshman, sophomore, upper division, and graduate). Among the dependent variables are scores on the seven items of a teaching performance scale. The items ask the student if

the instructor: 1) was well-prepared; 2) stimulated student thinking; 3) was actively helpful to students; 4) explained course objectives; 5) was fair and impartial in grading; 6) explained difficult material; and 7) if the students felt they learned a great deal. An average of the seven items is included as well.

A "Student Repeats Per Course" variable was calculated by counting the number of times each student in a particular class had previously been in a class with that same instructor, and dividing this total by the class size. For example, if only 2 students in a class of 20 had each had their current instructor for one other class, the value of the "Repeats" for that class would be .10. The "Repeats" variable is simply the mean of this "Repeats" measure for all classes of a particular course.

Results

The means of each item, of the average of the seven items, and of the "Repeats" are presented for each course in Table 1. F ratios and levels of significance for each dependent variable are also given.

For all courses combined, the item "Well-Prepared" had the highest rating (4.54), while the item "Explains Difficult Material" had the lowest (3.95). Significant differences ($p < .01$) over the twenty courses were found for all dependent measures except the items "Course Objectives Explained" and "Grades Fair and Impartial."

In Table 2 are presented the results of the factorial style analysis of variance by course type and course level. Courses for degree majors were rated significantly higher ($p < .05$) than service courses on all variables except "Well-Prepared," "Course Objectives Explained," and "Grades Fair and Impartial." The value of "Repeats" was .23 for service courses and 1.07 for major courses.

Table 1

Analysis of Variance by Course Effectiveness for Evaluations of Teaching for Twenty Selected Mathematics Courses

		Means										
Course #	Course Title	Well-Prepared	Stimulated Student Thinking	Actively Helpful to Students	Course Objectives Explained	Grades Fair/Impartial	Explains Difficult Material	Students Learn Great Deal	Average of Seven Items	Repeats	Number of Classes	Number of Students
Freshman												
102 (S)*	Intermediate Algebra	4.53	4.00	4.27	4.25	4.45	4.00	3.95	4.21	.08	89	1908
104 (S)	College Algebra	4.47	4.00	4.32	4.20	4.48	3.88	3.98	4.19	.14	70	1084
107 (S)	Elementary Statistics	4.59	4.15	4.49	4.29	4.57	4.17	4.08	4.34	.10	85	1452
126 (S)	Pre-Calculus	4.40	3.98	4.37	4.23	4.50	3.87	4.02	4.20	.19	69	1036
Sophomore												
211 (P)	Calculus I	4.67	4.20	4.37	4.32	4.45	4.02	4.21	4.32	.22	66	1100
212 (P)	Calculus II	4.45	3.99	4.17	4.16	4.38	3.79	4.07	4.14	.32	60	906
214 (P)	Calculus III	4.36	3.69	4.11	4.07	4.22	3.53	3.75	3.96	.50	17	211
215 (P)	Calculus IV	4.60	4.15	4.37	4.35	4.52	4.00	4.22	4.32	.36	27	327
Upper Division												
335 (M)	Matrix Theory	4.68	4.22	4.42	4.24	4.37	4.08	4.13	4.31	.61	16	208
435 (M)	Linear Algebra	4.71	4.24	4.49	4.19	4.53	4.00	4.22	4.34	1.08	15	128
447 (S)	Statistical Methods I	4.17	3.76	4.16	4.11	4.51	3.67	3.84	4.12	.59	12	161
448 (S)	Statistical Methods II	4.72	3.95	4.28	4.18	4.65	4.00	3.95	4.25	1.46	10	146
451 (M)	Mathematics Statistics I	4.53	3.85	4.25	4.01	4.29	3.55	3.66	4.02	.55	6	51
452 (M)	Mathematics Statistics II	4.71	4.29	4.56	4.23	4.72	3.97	4.20	4.38	1.92	5	27
455 (M)	Numerical Analysis I	4.80	4.52	4.37	4.36	4.45	4.14	4.30	4.42	1.46	7	52
456 (M)	Numerical Analysis II	4.87	4.55	4.63	4.47	4.60	4.42	4.48	4.57	1.88	6	44
Graduate												
795 (S)	Applied Statistics for the Behavioral Sciences I	4.33	4.07	4.35	4.14	4.58	3.75	4.09	4.19	.30	11	112
796 (S)	Applied Statistics for the Behavioral Sciences II	4.25	3.88	4.15	3.95	4.23	3.52	3.89	3.98	.56	13	131
820 (M)	Advanced Matrix Algebra	4.76	4.49	4.68	3.91	4.56	4.03	4.31	4.39	.86	4	36
856 (M)	Linear Statistical Models I	5.00	4.40	5.00	4.60	4.60	4.60	4.69	4.70	1.17	2	14
TOTAL		4.54	4.06	4.34	4.23	4.47	3.95	4.05	4.23	.32	590	9144
F Ratio		2.18	2.66	2.01	1.29	2.46	1.92	2.33	1.87	53.60		
p		.00	.00	.01	.18	.03	.01	.00	.01	.00		

* S - Service course; P - Predominantly service but includes some majors; M - Majors.

Table 2

Factorial Analysis of Variance by Type and Level
for Evaluations of Teaching Effectiveness for Twenty Selected Mathematics Courses

	Means									
	Well-Prepared	Stimulated Student Thinking	Actively Helpful to Students	Course Objectives Explained	Grades Fair/ Impartial	Explains Difficult Material	Students Learn Great Deal	Average of Seven Items	Repeats	Number of Classes
Course Type										
Service	4.52	4.03	4.32	4.23	4.47	3.94	4.04	4.22	.23	529
Major	4.72	4.28	4.48	4.23	4.48	4.05	4.19	4.35	1.07	61
Course Level										
Freshman	4.50	4.04	4.36	4.25	4.50	3.99	4.01	4.24	.12	313
Sophomore	4.55	4.07	4.27	4.20	4.42	3.89	4.11	4.22	.31	170
Upper Division	4.72	4.15	4.38	4.21	4.50	3.97	4.09	4.29	1.05	77
Graduate	4.40	4.07	4.35	4.05	4.43	3.75	4.07	4.16	.54	30
Type by Level										
Service by Freshman	4.50	4.04	4.36	4.25	4.50	3.99	4.01	4.24	.12	313
Service by Sophomore	4.55	4.07	4.27	4.24	4.42	3.89	4.11	4.22	.31	170
Service by Upper Division	4.74	3.85	4.22	4.14	4.57	3.81	3.89	4.18	.93	22
Service by Graduate	4.29	3.97	4.24	4.04	4.39	3.63	3.98	4.08	.44	24
Major by Upper Division	4.71	4.26	4.46	4.24	4.47	4.03	4.17	4.33	1.08	55
Major by Graduate	4.84	4.46	4.78	4.14	4.57	4.22	4.43	4.49	.96	6
F Ratios	F p	F p	F p	F p	F p	F p	F p	F p	F p	
Main	4.20 .00	4.37 .00	3.21 .01	1.52 .20	2.25 .06	2.25 .06	3.53 .01	1.83 .12	129.20 .00	
Type	1.01 .32	14.55 .00	8.34 .00	1.00 .32	.22 .64	3.91 .05	8.59 .00	5.09 .02	7.23 .01	
Level	2.19 .09	1.51 .21	2.24 .08	2.03 .11	2.57 .03	2.49 .06	2.87 .04	.37 .46	54.09 .00	
Interaction	5.59 .02	.09 .77	1.53 .22	.00 .97	2.75 .10	1.02 .31	.47 .49	1.26 .26	6.13 .01	
Explained	4.47 .00	3.52 .00	2.88 .01	1.22 .30	2.35 .04	2.00 .08	2.92 .01	1.71 .13	104.58 .00	

Upper division courses were rated the highest on most measures; they were not rated highest on "Course Objectives Explained," "Explains Difficult Material," and "Students Learn Great Deal." The value of "Average Repeats" for upper division courses was 1.05.

Graduate Courses were rated lowest for the following items: "Well-Prepared," "Course Objectives Explained," "Explains Difficult Material," and the average of the seven items. "Stimulated Student Thinking" and "Students Learn Great Deal" were rated lowest for freshmen courses. Significant differences ($p < .05$) between levels were found for "Grades Fair and Impartial," "Students Learn Great Deal," and "Average Repeats." Significant interaction effects were found for "Well-Prepared" ($p < .05$) and for the "Repeats" variable ($p < .01$).

A repeated measures type of analysis of variance was performed using the seven measures of teaching effectiveness and selected courses having an N of ten or more classes. An analysis for all twenty courses was also included. These results are presented in Table 3. Significant differences between items were found for each selected course as well as for all courses combined. Reliability coefficients were computed to provide on estimate of the level of consistency across the seven items. These reliability coefficients (Cronbach's alpha) ranged from .95 to .96, while the standardized item alpha coefficients ranged from .95 to .97. This degree of stability is considered high in terms of measurement applications (Stanley, 1971).

Conclusions & Implications

The purpose of this paper has been to investigate the influence of course and characteristics of the course on measures of teaching effectiveness. Data on 20 courses, 590 classes, and 9144 evaluations over a

Table 3

Repeat Measures Analysis of Variance
for Measures of Teaching Effectiveness
by Selected Courses and Total

Course #	Course Title	Number of Classes	F	p	Alpha	Standardized Item Alpha
102	Intermediate Algebra	89	86.09	.00	.95	.95
104	College Algebra	70	55.67	.00	.96	.97
107	Elementary Statistics	85	68.52	.00	.95	.96
126	Pre-Calculus	69	48.97	.00	.96	.97
211	Calculus 1	66	58.82	.00	.96	.97
212	Calculus 2	60	48.36	.00	.95	.95
All 20 Courses		590	433.03	.00	.95	.95

four-year period were used in several analysis of variance procedures.

Differences between courses were found for five of the seven items as well as for the average of the seven items. These were "Well-Prepared," "Stimulated Student Thinking," "Actively Helpful to Students," "Explains Difficult Material" and "Students Learn Great Deal." As would be expected, there were differences between courses for the "Repeats" measure, with higher values observed for upper division courses and for graduate courses for degree majors.

Differences between types of course were found for four items, the average of the items, and the "Repeats." Differences between levels were found for only two items and "Repeats." Repeated measures of variance yielded relatively high coefficients of reliability as well as significant differences between the seven items.

Previous research has shown course variables to affect student ratings of instructor performance in college level courses in general. The results of this study indicate the importance of taking into account the particular course, its type, and its level when making use of student evaluations of teaching performance in mathematics courses.

References

Abrami, P.C., Dickens, W.J., Perry, R.P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? Journal of Educational Psychology, 72, (1), 107-118.

Abrami, P.C., Perry, R.P., & Levanthan, L. (1982). The relationship between student personality characteristics, teacher ratings, and student achievement. Journal of Educational Psychology, 74, (1), 11-125.

Braskamp, L.A., Ory, J.C., & Pieper, D.M. (1980). Student written comments: Dimensions of instructional quality. Journal of Educational Psychology, 73, (1), 65-70.

Dowell, D.A., & Neal, J.A. (1982). A selective review of the validity of student ratings of teaching. Journal of Higher Education, 53, (1), 51-62.

DuCette, J., & Kenny, J. (1982). Do grading standards affect student evaluations of teaching? some new evidence on an old question. Journal of Educational Psychology, 72, (3), 308-314.

Greene, J., Jr., Prather, J.E., & Sturgeon, J.E. (1983, May). Using administrative data as unobtrusive indicators of teaching performance. Paper presented at The Association for Institutional Research Forum, Toronto, Ontario, Canada.

Hills, J.R. (1974). On the use of student ratings of faculty in determination of pay, promotion and tenure. Research in Higher Education, 2, (4), 317-324.

Hofman, J.E., & Kremer, L. (1980). Attitudes toward higher education and course evaluation. Journal of Educational Psychology, 72, (5), 610-617.

Howard, G.S., & Maxwell, S.E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? Journal of Educational Psychology, 72, (6), 810-820.

Howard, G.S., & Maxwell, S.E. (1982). Do grades contaminate student evaluations of instruction? Research in Higher Education, 16, (2), 175-188.

Marsh, H.W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 17, (1), 219-237.

Marsh, H.W. (1982). Validity of students' evaluations of college teaching: A multitrait, multimethod analysis. Journal of Educational Psychology, 74, (2), 264-279.

Overall, J.U., & Marsh, H.W. (1980, April). Relative influence of course level, course type, and instructor on students' evaluation of instruction. Paper presented at the meeting of the American Educational Research Association, Boston. (Eric Document Reproduction Service No.ED 187 176)

Prather, J.E., Massey, F.A., & Greene, J., Jr. (1983, August). Students' evaluations of instructors in introductory statistics courses. Paper presented at the Joint Statistical Meetings, Toronto, Ontario, Canada.

Prather, J.E., Massey, F.A., Greene, J., Jr., & Sturgeon, J.S. (1984, January). Evaluations by students in mathematics courses of the effectiveness of teaching. Paper presented at the meeting of the American Mathematical Society, Louisville, KY.

Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), Educational Measurement (pp. 356-442). Washington, DC: American Council on Education.